Haile Quals Day 1

Multiple "types" of memory (e.g., declarative, working, etc.) have been proposed. To what extent does the neurophysiological evidence support or contradict the idea of distinct memory systems as opposed to a more 'unified' memory system that is flexible enough to operate differently depending on the task.

The best way to answer this question is by first asking, what is memory for? A learning agent needs to gather information about its surroundings so it can survive and thrive in it. Some moments in the agent's environment call for rapid learning and flexibility, others call for precision and speed (Hikosaka et al. 2013); and, in a dynamic environment, some learned information is no longer needed, and others might be more valuable (Anderson, 2000). This push and pull between flexibility, reliability, and readiness, suggests multiple systems for memory. However, most models of multiple memory systems do not capture the complexity of the networks of brain regions that give rise to memory function, and we should not surmise that a type of memory is supported by individual brain regions. That being said, I argue that the neurophysiological evidence supports the idea of multiple memory systems because we can demonstrate that different brain networks compute different types of information, some of these networks function in parallel, and loss of critical network components results in predictable behavioral deficits in memory function. But it cannot be ignored that we, as human researchers, prefer to organize complex systems using multiple, discretized simpler components, and even the neurophysiological data might show that bias (Keren and Schul, 2009; Buszáki, 2020). Additionally, it can be argued that some brain regions belong to multiple networks and represent multiple types of information, so much so that we could argue for a unitary system of memory made up of domain-general and domain-specific sub-systems.

The idea of a multi-component memory systems is prevalent, and we often name and examine distinct brain regions when we talk about individual memory systems. These distinct components of memory are characterized by the time course within which they acquire information (e.g., complex motor skills require several hours of practice), how long the information lasts (e.g., briefly rehearsed information is lost as soon as rehearsal ends), and the type of information that is learned (e.g., patterns of motor movements in a skill compared to what you had for dinner yesterday). Conventionally, these are divided into the more short-term working memory and the several types of long-term memory: episodic, declarative, and procedural memory. To briefly provide over-simplified definitions, working memory is a function that enables a learner to attend to and manipulate currently relevant information (Miller, 2013), declarative long term memory stores nameable and verbally describable information, episodic memory stores sensory information that are tied to specific events in the agent's experience and, lastly, procedural memory stores complex motor patterns that are demonstrable in a skill and habits.

There are several reasons why we have prominent ideas of multi-component memory systems. Firstly, most of our understanding about memory comes from observable behavior, and perhaps, introspection. A multi-component system makes sense because we can characterize a type of memory that is engaged for a specific situation and the level of experience of the agent. We can easily measure capacity limits for items we can hold in mind, demonstrate skill memory that is distinguishable from mere items in a list, and how long it takes for these types of memory to be formed and lost. For instance, we do not verbally rehearse motor rules required for walking because such a well-trained skill is represented as non-declarative motor patterns, but a new driver might verbally remind themselves what to do when traffic lights turn red. Secondly, we normally dissect and discretize complex systems to make them easier to understand and investigate more deeply resulting in multi-component systems (Keren and Schul, 2009). For instance, Baddeley's early fascination with verbal information led to an influential model of short-term memory that was primarily a brief phonological store, with an executive overlord that decided what items to maintain in this store. But Baddeley's experience and experiment dictated that he add to his model of working memory more sub-systems that represented other types of information like visuo-spatial and episodic information (Baddeley, 2012).

Thirdly, influential experiments with brain lesion or, other specific dysfunctions showed that damage to very specific brain regions resulted in behavioral deficits only in some memory types and not others. There is no physiological evidence that suggests unitary regions that compute the above-mentioned gross memory systems, but these are very valuable pieces of physiological evidence that show, at least that there is some "division of labor" regardless of whether we view memory as a unified system or one with multiple components. Let us discuss the role of the hippocampus for instance.

We highlight the hippocampus when talking about episodic or declarative memory because damage in the hippocampus had resulted in inability to form new declarative or episodic memories, but procedural long-term memory function is nearly intact (knowlton et al. 1994). The famous case of Henry Molaison, patient HM, who had had bilateral removal of the hippocampus, amygdala and surrounding tissue, had major deficits in making new memories and recalling some past episodic memories (Squire, 2009). But Molaison's ability to learn new procedural motor skills is intact, he'll just likely not remember that he learned them. Similarly, Knowlton et al., 1994, have shown that amnesic patients with damage to the hippocampus or other related structures in the limbic system, like mammillary bodies and thalamus were able to learn new procedural responses. These patients acquired the correct responses, 'Rain' or 'Sun' associated with a set of specific combination of 4 images that contained basic shapes (circle, square, diamond etc.) in a task called the Weather Prediction Task (note that there is a complication here that will be discussed later in this essay).

The hippocampus in necessary for coding new information to cortex and for recall of past information. But it is often taken for granted that the hippocampus is only one region of the many that are involved in the acquisition, processing and storage of the bits that make up both

episodic and declarative memory. There is evidence that shows, depending on the type of information, memory traces in long term memory are stored in distributed brain regions that are responsible for the specific information's integration and representation. For example, memories for objects are stored in regions that process objects, faces in face-selective regions, words in language areas etc., all outside the hippocampus. Studies with non-human primates and other animals have shown that memory formation is sensitive to rewarding events and is partial to the features of stimuli that have high value. This suggests that information is processed by the concerted effort of additional regions like the prefrontal cortex and basal ganglia (McDonald and Hong, 2013; McNab and Klingberg, 2008). The same is true for other types of memory.

Procedural long-memory, as suggested above, is always contrasted to declarative long-term memory because it involves learning intricate, complex motor patterns in response to stimuli that are reinforced by rewards. While we can demonstrate that non-declarative information is learned through specific centers in the basal ganglia special care must be taken since the basal ganglia have far reaching effects. To start, learning through rewards occurs via dopamine signaling that is initiated in the basal ganglia— dopamine neurons in the Ventral Tegmental area are sensitive to rewards and specifically signal that a reward is imminent given some response by the learner after some training (Schultz et al., 1997). The basal ganglia, especially the dorsal striatum directly and indirectly send and receive signals to and from the whole cortex (e.g., the prefrontal cortex, the thalamus, and the hippocampus; Packard and Knowlton, 2002). Disruption in the basal ganglia, as in the case of Parkinson's Disease, where there is damage to the dopamine neurons in Substantia Nigra, or lesions results not only in motor learning and planning deficits but also category and habit learning deficits (Shohamy et al., 2008; Packard and Knowlton, 2002). The basal ganglia are a vastly important part of learning and memory, but hippocampal learning of categories and associations is possible even when there is damage to the basal ganglia (Poldrack et al, 2001). This suggests that we have distinctly dissociable memory systems since the entire system of learning and information storage is not lost.

Further evidence for a multi-component memory system is provided by the fact hippocampus-centric and striatal learning of stimulus-response associations seem to occur in parallel and compete for control of behavior (Poldrack et al., 2001; Poldrack and Packard, 2002; McDonald and Hong, 2013). In an influential fMRI study, Poldrack et al., 2001, showed that the medial temporal lobe, which contains the hippocampus, is active in the beginning of learning a task like the Weather Prediction Task, while the basal ganglia is deactivated. But as the subject encounters more learning trials, activity in the caudate part of the basal ganglia increases along with a decrease in the activity of the medial temporal lobe. This negative correlation in the activity of these two regions suggest that they actively compete for control of behavior. There are numerous examples of this difference in early and late representation of knowledge. Anderson describes a model of skill learning where initial declarative representations of how to perform a task are *proceduralized* into more efficient and fast motor procedures (Anderson, 1982; Tenison et al.,

2016). For example, passwords we type very frequently are coded as motor patterns of key-presses and are sometimes challenging to recall (alternatively, how quickly can you recite only the last 4 digits of your phone number?).

Even though the inclusion of some brain regions in memory networks are critical for specific memory types, regions like the basal ganglia and prefrontal cortex are part of multiple networks, suggesting a more unified memory system. If we go back to a more fundamental explanation for what memory is for, we could argue that the physiological evidence points to a hierarchically organized system where some networks, especially those involving the prefrontal cortex, build rich, multi-modal, representations of a learner's environment along with the most efficient and relevant behavioral responses. Taking the example of complex categories, Miller, 2013, explains that declarative and non-declarative systems provide information that the prefrontal cortex assembles into more complex knowledge. I have not seen any papers that argue this, but how much of our conscious experience of our knowledge or skill is resultant from only declarative or only procedural memory?

A final important question we should ask is, how much is our point of view of multiple and discrete memory types more due to a bias in how we have historically asked questions and organized information, even those studies that include physiological evidence? In an argument against 2-component processing systems, Keren and Schul, 2009, argue that most empirical evidence for the existence of such a division in processing systems are not conclusive. They do not argue for a multi-component or even a single-component processing system but point out that characterizations of higher cognition are not well tested empirically and are ill defined. They suggest that there is large appeal for breaking up complex systems into sub-components because it lends some simplicity. I believe there is some truth to this as we often struggle to define, very clearly, the differences between, for example, implicit and explicit knowledge or procedural and declarative information. Similarly, Buszáki, 2020, argues that a lot of research attempt to find brain correlates of already existing psychological phenomena with correlational studies. He cautions against this, as have others, and suggests an 'inside-out' approach, where clearly defined and measured brain phenomena should guide the definition of the function, they give rise to.

In summary, physiological evidence involving, lesion and patient studies, as well as well controlled brain imaging studies point more to a multi-component view of memory. But we should be cautious as we tend to go look for evidence in the brain for prominent psychological theories and phenomena that may have historical biases.